TITLE: Robust Head Tracking Based on Hybrid Color Histogram and Random Walk Kalman Filter

Robust Head Tracking Based on Hybrid Color Histogram and Random Walk Kalman Filter

Gwang-Myung Kim[1], DongCheng Lin[2], Jung H. Kim[1], Sung H. Yoon[2]

[1]Department of Electrical and Computer Engineering
[2]Department of Computer Science
North Carolina A&T State University
Greensboro, NC 27411
Tel: (336) 334-7760 x 219 Fax: (336)334-7244

Abstract

This paper examines a new robust color scheme and an adaptive object tracking technique. There are several popular color schemes used in face tracking which include Normalized RGB, Hue, Saturation, and Hybrid type of colors. Hybrid color schemes provide improved results compared to any single color scheme technique. Extensive experiments show the new robust Hybrid color scheme produced superior results in various lighting conditions. In conjunction with the robust hybrid color scheme to track head movements a supporting algorithm was needed to approximate the random path of the head movement. Kalman filter is a famous estimation technique in many areas to predict the route of moving object. We tested and developed a random-walk Kalman filter to track unpredictable and fast moving objects. The random-walk Kalman filter tolerates for tracking of quick random movements made by a person, which was not accommodated by linear tracking techniques.

1. Introduction

For many computer vision applications, such as automatic speech recognition, 3D animation, and surveillance a robust and reliable automatic head tracking technique in various unmodified environments is vital. Recent research in this area shows great progress and promise. There are many approaches to track the head position on an image sequence. Some tracking modules are based on feature invariant, which is used to find out a structural feature, some are based on template matching, which is using a stored pattern to track head position (pattern can be 2D or 3D). Others include appearance-based method, which is using a trained model from a set of images to capture the representative variability of facial appearance. In this paper we explore a combination of a hybrid color scheme module and a random-walk Kalman filter to track random head movement in a variety of environments.

Many researchers have exploited the relative uniqueness of skin color to track faces. Human skin color has been used and proven to be an effective feature in many applications. A weakness of these systems is their heavy reliance upon skin color that forbids skin-colored objects in the background and, more importantly, forbids the subject from turning around so that the back of his head, rather than this face, is visible [1].

Color image histogram is an effective method for the purpose of object recognition, segmentation or tracking. Color histograms are relatively invariant to many complicated, non-rigid motions like translation, rotation about the imaging axis, small off-axis rotations, scale changes and partial occlusion. The color histogram percentile features are useful to recognize the pattern of human face with relatively low complexity. Many methods have been proposed to build a skin color model. In this paper we proposed a new Hybrid color scheme with the support of additional Hue and Saturation analysis features that provide noticeable improvement in performance in various lighting conditions.

The Kalman filter is an optimal estimator to predict the next position of a moving object. It addresses the general problem of trying to estimate parameters of interest from indirect, inaccurate and uncertain measurements. However, general purpose of Kalman filter is only working well under slight movement and gradual speed on the image sequence. We need adaptive methods to overcome this problem.

Section 2 will cover the color performance analysis in head tracking to show the improved result of our new color scheme compared to result of other systems. Section 3 covers random-walk Kalman filter to trace correct location of unpredicted and rapidly moving object. Finally, section 4 will provide conclusion of experiment result.

2. Analysis of Color Scheme for Head Tracking

In the RGB model, a color is expressed in terms that define the amounts of Red, Green and Blue light it contains. Normalized color space is a popular color representation to specify human skin color patterns. Since under normal lighting conditions the brightness of the face is not important for characterizing skin colors, we can represent skin-color in the chromatic color space. Chromatic colors, known as "pure" colors in the absence of brightness, are defined by a normalization process [2].

$$C_r = R / (R + G + B)$$
$$C_b = B / (R + G + B)$$

Even though the most common way of representing color is through the RGB color space. In this paper we can see this color model is quite sensitive to lighting conditions since the color attribute is combined with the brightness. Hue (color) component can be used for facial region localization because it is comparatively insensitive to illumination changes. Hue image is obtained by logarithmic color-space transform, RGB to HSV. However, simple Hue image can be easily affected by complex background texture. Additional Saturation component can compensate this lack of robustness to the intricate environment.

S. Birchfield [2] introduced his own color scheme; in our experiments we call it *the Stanford scheme*, which uses color space consisting of scaled versions of the three axes *B-G*, *G-R*, and *B+G+R*. The first two contain the chrominance information and are sampled into eight bins each, while the last one contains the luminance information and is sampled more coarsely into four bins. The big difference in his method is that he also considers luminance information. By using this scheme we could get fairly good tracking result. However, this scheme shows partial dependency on light condition and background texture.

We attempted to find a new color scheme that is robust enough for various light and background conditions. From our previous experiment, *Stanford scheme* showed a better result compared to other methods. But in addition to this scheme, the characteristic of insensitivity to illumination is required for a practical and dependable tracking module. A new Hybrid color scheme that utilizes additional Hue and Saturation features is the one we chose to achieve this goal.

The research was executed with various sequences of images under different light condition, background, and persons. For the objective comparison of result, all of four sequences were obtained from Vision lab website of Stanford University. Person in a sequence is always inside of frame by controlling the camera movement. These sequences include different races, light condition and background. Importantly, linear prediction technique was exploited to predict next head position for this test.

Table1 and 2 shows head tracking result of various color schemes we chose for test. As it is shown below, Hybrid color histogram with (20(Stanford) + 4(Hue) +4(Saturation)) bins gives the best results compared to Hue (16), Hue and Saturation (8 + 8), Normalized RGB, *Stanford* scheme (20) and Hue-hybrid (20 + 8(Hue)) color histogram.

We employed the average distance from the true center (Table 1) and the average success rate (Table 2) as performance measurements. True center of each frame was firstly obtained by manual operation through the whole sequence. Average distance was calculated based on this series of true center points. Each test was implemented both of *X* and *Y* directions to provide a better benchmark of tracking result evaluation.

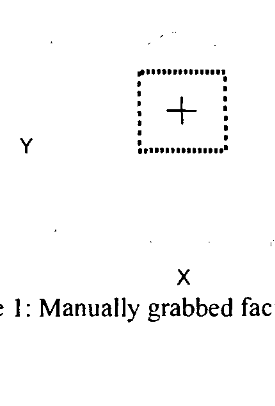


Figure 1: Manually grabbed facial region

Figure 1 shows the facial area and center point of that region. Hit number for each sequence of Table 2 is counted up when the destination point is located inside of this rectangular region. There is acceptable error range of five to ten pixels depends on the image.

From the result of Table 1 and Table 2, Hybrid color (20+4+4) gives 5.86 pixels distance to the X axis and 8.96 pixels to the Y axis. This is fairly good result compared to other two competent color schemes of Hybrid (20+8) and Stanford's (20). The result of Table 2 well supports this consequence.

We can expect better result only with additional Hue color (20+8). However, this color gave worse result for the sequence 3. Success ratio to the Y axis of sequence 3 is less than 50%. This means that Hue information is not stable enough to support Stanford color completely.

Stanford color scheme includes Normalized color and Regular RGB color. Even though their color system provides comparatively good results, it is still not robust enough under different conditions. Our test result shows that additional Hue and Saturation color features can attenuate the performance limitation of Stanford color.

3. Random-walk Kalman Filter

A robust head tracking requires a reliable prediction module for the estimation of the of the random moving objects. Our approach is on the base of Stan Birchfield's [2] method, which using intensity gradients, color histograms, and simple linear prediction. In gradient, an ellipse template is used to calculate the total gradient value around this ellipse within a suitable search window and then acquires a maximum value. In color, a face color histogram model will be created and used to match within the above search window. Birchfield also used a linear prediction to predict the search window on the oncoming frame according to the position of the previous 2 frames.

The main problem of the Birchfield method is the lack of accuracy if the moving speed of the head is too fast or the frame rate is too low. The result is a unreliable prediction window and the head position will be distracted. In this case, the way to improve the tracking performance is to increase the search range of search window, however this will cause the processing speed down. So, there exists a limitation in using the linear prediction algorithms used by Birchfield.

In order to overcome this problem, we propose a random walk Kalman filter to predict the search window with a center of head position and a suitable range on the consecutive frames, and then update this prediction using the measurement value of the tracking head.

Kalman filter is an optimal estimator. It addresses the general problem of trying to estimate parameters of interest from indirect, inaccurate and uncertain measurements. Due to its recursion, new measurement data can be fed back to system as they arrive, so it can be used in real-time image processing system.

Kalman filter estimates a process by using a form of feedback control: the filter estimates the process state at some time and then obtains feedback in the form of (noisy)measurements. As such, the equations for the Kalman filter fall into two groups: time update equations and measurement update equations [4]. The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the *a priori* estimates for the next time step. The measurement update equations are responsible for the feedback–i.e. for incorporating a new measurement into the *a priori* estimate to obtain an improved *a posteriori* estimate. To adapt this prediction method to our random tracking needs we introduce new algorithms.

In our system, we construct the system model as random walk. Some related equations are as follows:

The state vector $\hat{x}_k = [x_k, y_k]$, where x_k, y_k indicate the center position of head on the kth frame image.

The measurement vector $z_k = [x_{z_k}, y_{z_k}]$, where x_{zk}, y_{zk} express the measurement value from our approach.

$$(1) \quad \hat{x}_k^- = u(t),$$

u(t) = unity Gaussian white noise, that is random walk which means it has zero mean and unity variance [3].

$$(2) \quad z_k = Hx_k + v_k$$

From (1), (2), we can construct parameters of Kalman filter as follow:
Transmition matrix

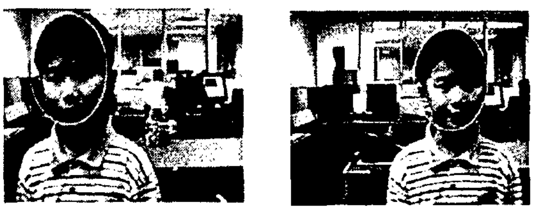
$$A=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, H=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$

$$Q=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, R=\begin{bmatrix}0.1 & 0\\ 0 & 0.1\end{bmatrix},$$

The initial *a priori* estimate error

$$P_0^- =\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

It will show different performance by using different frame rate sequence of image. We captured some different image sequence with different frame rate, 10,24 frames per second. If we use 24 fps image sequence, there are no problems. Following sample results are from a 10 fps image sequence. In this sequence, the maximum head displacement between 2 consecutive frames is about 62 pixels. If using the linear prediction, the center of search window on the next frame would be out of tracking, particularly on turnover motion. That means it can't get the good result. However, we got good results in our approach using random walk Kalman filter. Figure 2 (a) and (b) show our experiment result of head tracking by using random walk Kalman filter.



(a) (b)

Figure 2 : Sample results from a 10 fps image sequence

Figure 3 shows the x-coordinate comparison of head position of Kalman filter, Birchfield's, and true center. The real head positions are recorded manually. There are several pixels calibration between Kalman filter and Birchfield's approach.

4. Conclusion

This paper presents a robust automatic visual tracking module that utilizes a new Hybrid color scheme with hue and saturation support and random-walk Kalman filter for the prediction of the head. From our test result, we can conclude that proper mixture of all of RGB, chromatic color, Hue, and Saturation gives the best result compared with other currently available color schemes to track the human face. Moreover, if it can be combined with random-walk Kalman filter, the resulting module should provide a robust and reliable tracking method that overcomes many current problems in predicting the correct position of random and fast moving objects. The improvements in these two modules shows great promise for the development of a robust head tracking for ASR and other computer vision applications.

References

[1] J. Yang, A. Waibel. A real-time face tracker. In *Proc. of WACV'96,* 1996, pp. 142-147.

[2] Stan Birchfield. Elliptical Head Tracking Using Intensity Gradients and Color Histograms, In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition,* Santa Barbara, California, pages 232-237, June 1998.

[3] Robert G. Brown and Patrick Y. C. Hwang, "Introduction to Random Signals and Applied Kalman Filtering", Second Edition, John Wiley & Sons, WC, p273-274, 1992.

[4] "Open Source Computer Vision Library Reference Manual", Intel Corporation, Chapter 19, Copyright © 1999-2001

Table 1 : Average Distance from the True Center (unit : pixel)

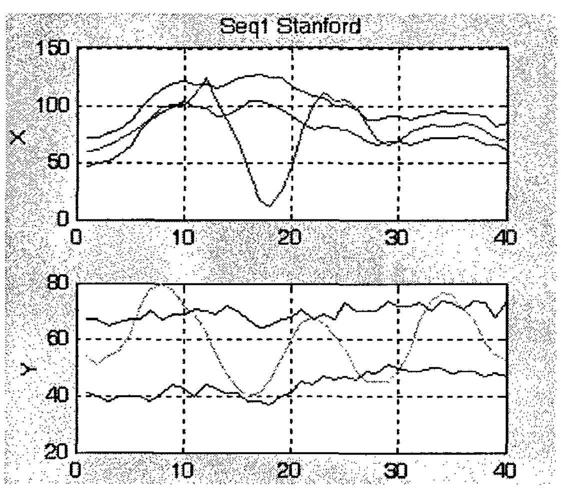
	Seq. 1		Seq. 2		Seq. 3		Seq. 4		Avg. (pixel)	
Hybrid (20+4+4)	5.49	8.49	7.44	5.98	5.31	9.9	5.19	11.46	5.86	8.96
Hybrid (20+8)	4.49	8.49	8.5	7.03	16.09	17.45	3.38	8.17	8.12	10.29
Stanford	16.99	9.61	8.72	11.49	3.52	10.08	3.4	7.82	8.16	9.75
Hue+Saturation	23.86	21.51	15.05	14.28	3.15	9.56	6.44	9.13	12.13	13.62
Hue	33.56	20.29	13.7	12.31	9.3	10.88	7.65	10.18	16.05	13.42
Normalized	25.21	9.89	34.13	36.8	30.83	17.82	4.85	10.97	23.76	18.87
	X	Y	X	Y	X	Y	X	Y	X	Y

X : x direction tracking result
Y : y direction tracking result

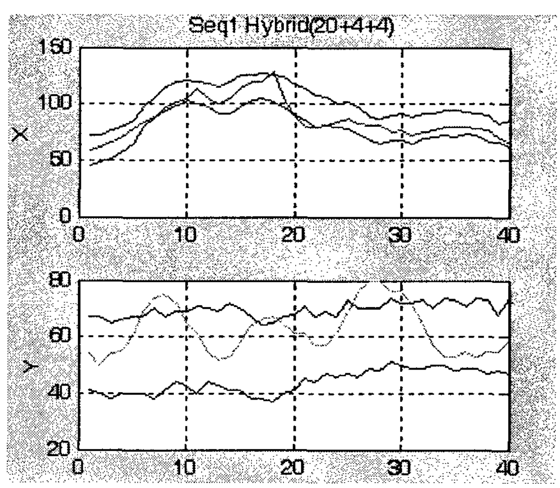
Table2 : Average Success Rate (Possibility to stay in the facial region through the whole sequence)

	Seq. 1 (40*)		Seq. 2 (65)		Seq. 3 (85)		Seq. 4 (101)		Avg. (%)	
Hybrid (20+4+4)	37	29	59	61	80	61	93	77	92.4	78.4
Hybrid (20+8)	39	33	51	59	59	40	101	97	85.9	78.7
Stanford	25	27	47	49	82	56	101	98	87.6	79.0
Hue+Saturation	14	18	36	35	81	62	91	81	76.3	67.4
Hue	14	21	33	39	62	49	84	80	66.3	64.9
Normalized	19	27	20	17	46	30	94	85	61.5	54.6
	X	Y	X	Y	X	Y	X	Y	X	Y

* : # of frames in a video sequence



(a)



(b)

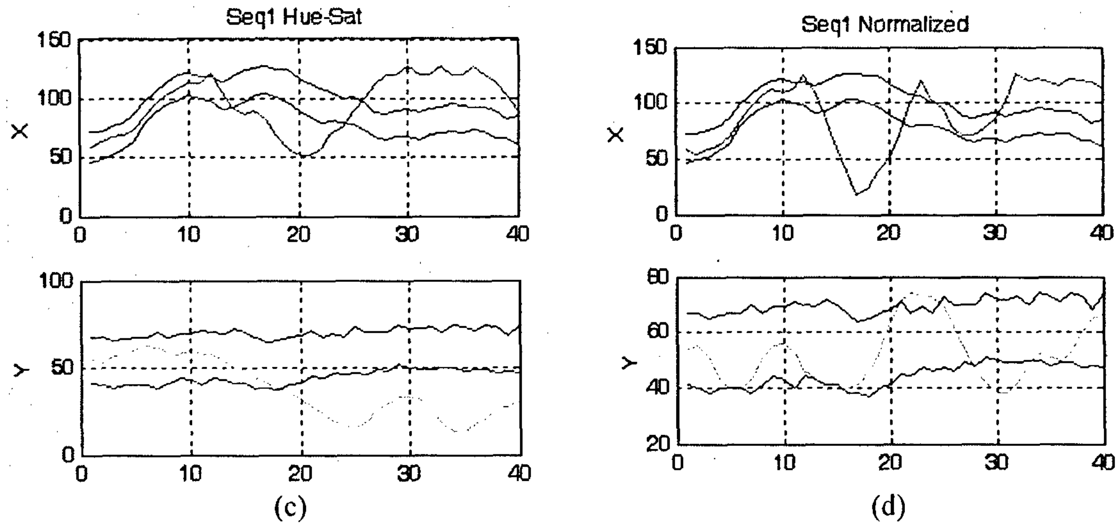


Figure 1 : (a) Stanford (B-G)+(G-R)+(R+G+B/3) (b) Stanford + Hue(4) + Saturation(4) (c) Hue + Saturation Color Scheme (d) Normalized Color

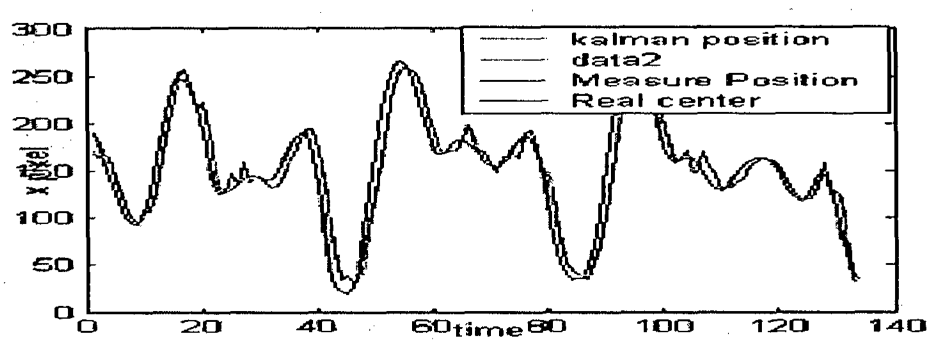


Figure 3 : Comparison x-coordinates of head position with Kalman filter, Birchfield, and real center position (manually recorded).